

Ontario Genomics Institute

2011 Progress and Applications of Next Generation Sequencing Workshop

Jan 20, 2011

Dennis McCormac

Ontario Genomics Institute/The Centre for Applied Genomics



Ontario **Genomics** Institute

OGI Mission



Use world-class research to create strategic genomics resources ...

Through research program management:

attract & support investment in Ontario's genomics research.

Through business development:

catalyze access to & impact of genomics capacity & resources created.

Through outreach activities:

raise visibility of genomics as well as its impact & associated issues.

... and accelerate Ontario's development of a globally-competitive life sciences sector.

Fosters access to Genomics Resources across Ontario



• Technology Services

- Genome Canada funds six S&T Innovation across Canada (TCAG)
- OGI has several affiliate platforms with services that are complementary to those of Genome Canada funded platforms
- Other OMICs platforms/facilities in the province



The Centre for Applied Genomics (TCAG)

Core facilities



1. DNA Sequencing & Synthesis

- DNA sequencing, NGS (1 GAI, 5 SOLiDs, 1 454) oligos

2. Microarray & Genotyping

- microarrays for genetic analysis, gene expression analysis

3. Cytogenomics

- Enhancing services for multi-species structural variation (including CNV) analyses using karyotyping, FISH mapping and targeted genotyping (expanded into ES and induced pluripotent stem (iPS) cells)
- *Database of Genomic Variants.*

4. Biobanking

- preparation of permanent cell stocks and DNA from blood or skin
- >50,000 cell lines banked

5. Informatics and Biostatistics

- has grown 30% in the past year and supports all other cores
- focus on high-throughput sequence analysis, genetic pathway analysis, high-performance and cloud computing, and training.



OGI Platform Affiliate members



- AAC Genomics Facility – Guelph
- London Regional Genomics Centre – London
- UHN Microarray Centre – Toronto
- Analytical Genomics Technology Centre (AGTC) – Toronto
- The Centre for the Analysis of Genome Evolution and Function (CAGEF) – U of Toronto
- StemCore Laboratories – Ottawa
 - CCBR – Toronto
 - OICR – Toronto

• *For more information about affiliates and this program visit:*

• <http://www.ontariogenomics.ca/business-development/ontario-genomics-platform-affiliates-program>



Technology Days



To increase visibility and use of genomics resources that Ontario researchers have helped develop

- A webinar/web conference co-hosted by OGI and the Ontario Institute for Cancer Research (OICR) that offered information about the Reactome Database - an open-source, manually-curated resource of core pathways and reactions in human biology
- A webinar/web conference co-hosted by OGI and The Centre for Applied Genomics (TCAG) about the Database of Genomic Variants (DGV), a curated database that catalogues structural variation in the human genome
- Two hands-on demos of bioinformatics tools for 'Pathway and Network Analysis' including Cytoscape and GeneMANIA, co-hosted by OGI and TCAG in Toronto
- A series of webinars/web conferences co-hosted by OGI and TCAG that offered information about mouse mutagenesis resources available through the NorCOMM project
- iREFWeb, Savant (genome browser),

OGI can help you highlight your resource to a broader audience



OGI Programs to enhance genomics resources



Platform Affiliates

- Genomics Technology Seeding (GTS) program

Workshops

- symposia on emerging areas of commercializable science
- vendor tutorials
- multi-vendor technology overviews

Genomics tech seeding project



To develop diagnostic testing for multigenic disorders using Next Generation Sequencing with targeted enrichment technologies



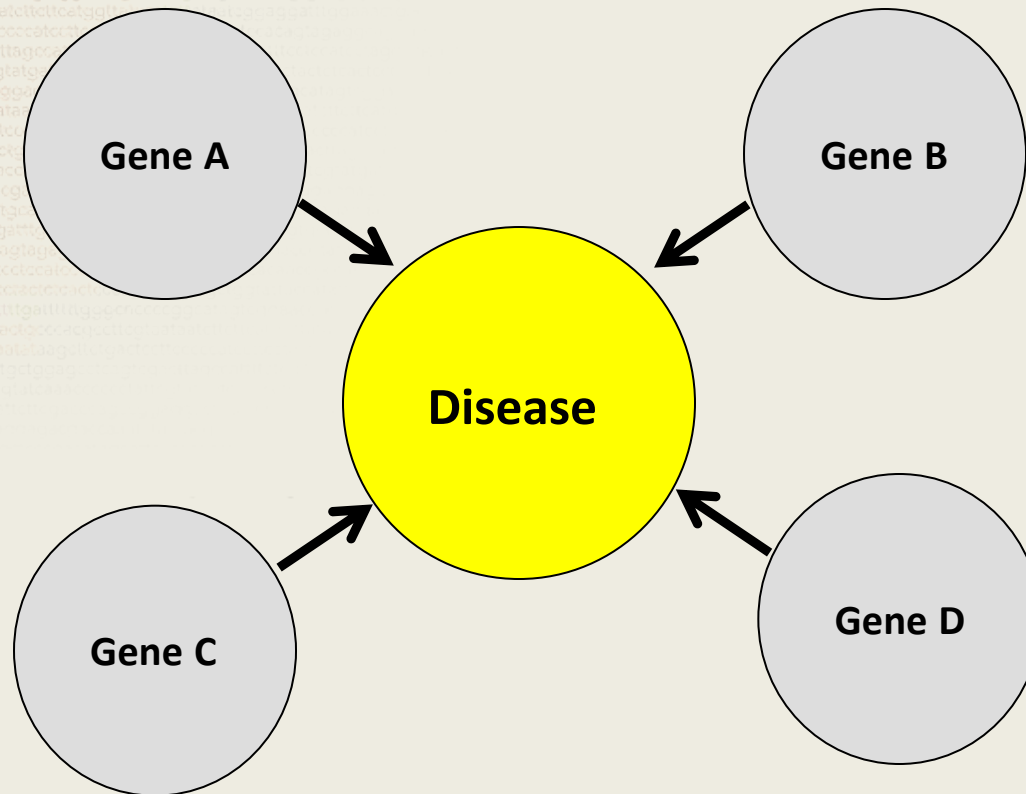
Targeted Resequencing for Clinical Diagnostic Testing (TCAG – DPLM)



- **Founded 1987**
- **The largest DNA diagnostic lab in Canada**
- **30 % of all DNA testing in Ontario**
- **28,000 samples per year**
- **135 genetic tests**
- **23% based on whole gene sequencing**

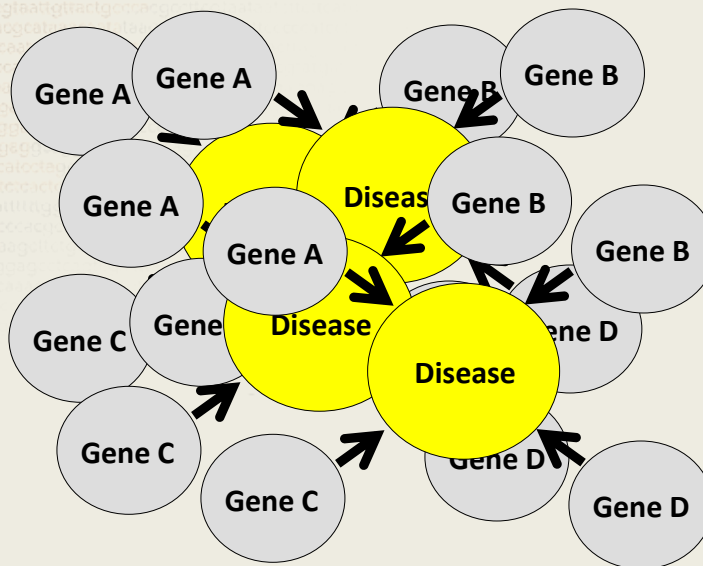


Complex Disease Heterogenetic Inheritance



Large number of rare variants with large effect

Complex Disease Phenocopies



Neuromuscular panel



Disease	Genes	# Exons	
HSP (AR)	SPG7	17	
	SPG11	40	
	SACS	10	
	SPG5A	6	
	SPG15	42	
	PNPLA6	35	
	SPG20	9	
	SPG21	9	
	CCT5	11	
	HSP (AD)	SPG3A	14
		SPG4	17
SPG6		5	
SPG8		29	
SPG10		29	
SPG13		12	
SPG17		12	
SPG31		7	
SPG33		13	
HSP (X-linked)		L1CAM	28
		SPG2	7
CMD	POMGNT1	22	
	FKTN	11	
	SEPN1	13	
	ITGA7	25	
	LAMA2	65	
	LARGE	16	
	POMT2	21	
	POMT1	20	
	FKRP	4	
	COL6A1	35	
	COL6A2	28	
	COL6A3	44	
	CAPN3	24	
	SGCA	10	
	SGCB	6	
SGCD	9		
SGCG	8		
DYSF	55		
TCAP	2		
TRIM32	2		
LGMD (AD)	LMNA	12	
	MYOT	10	
EDMD	CAV3	3	
	LMNA		
Total	EMD	6	
	44	803	



Essential Considerations



- **Sequence coverage** - 100% of all exons
- **Sequence Accuracy** - 100%
- **Sensitivity** - detect heterozygotes
- **Ease of use** - routine use by lab technologists with routine protocols
- **Cost per patient** - <\$3000 / patient
- **TAT** - 4 weeks

Cost Prediction



- **Sequence target**

\$400

- **Library prep**

\$1700

- **Sequencing**

\$300

- **Total**

\$2400

Targeted enrichment platforms



RainDance sequence enrichment

- Microdroplet PCR
- individual primer pairs/ microdroplet

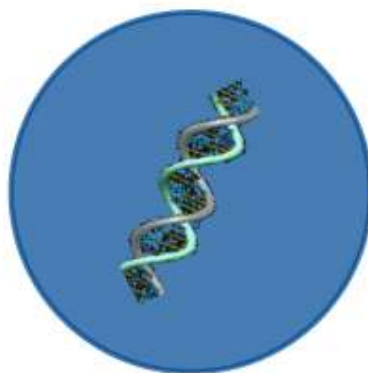
SureSelect (Agilent)

- Sequence capture
- Custom biotinylated RNA “baits” in solution

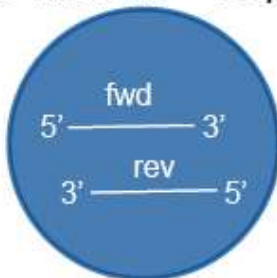
Microdroplet PCR (RainDance)



Inputs



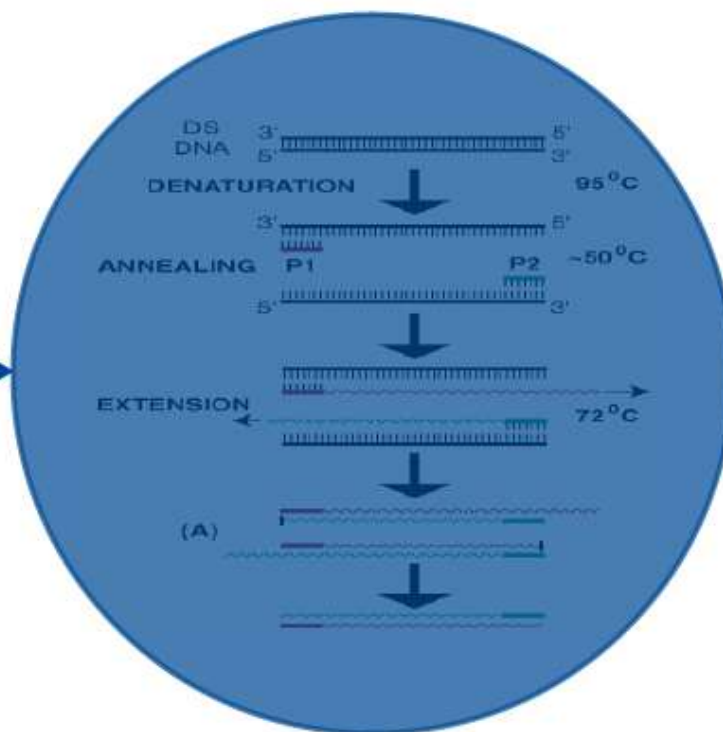
Genomic DNA Template Droplet
33 microns - 18 pL



Primer Pair Droplet
25 microns - 8 pL

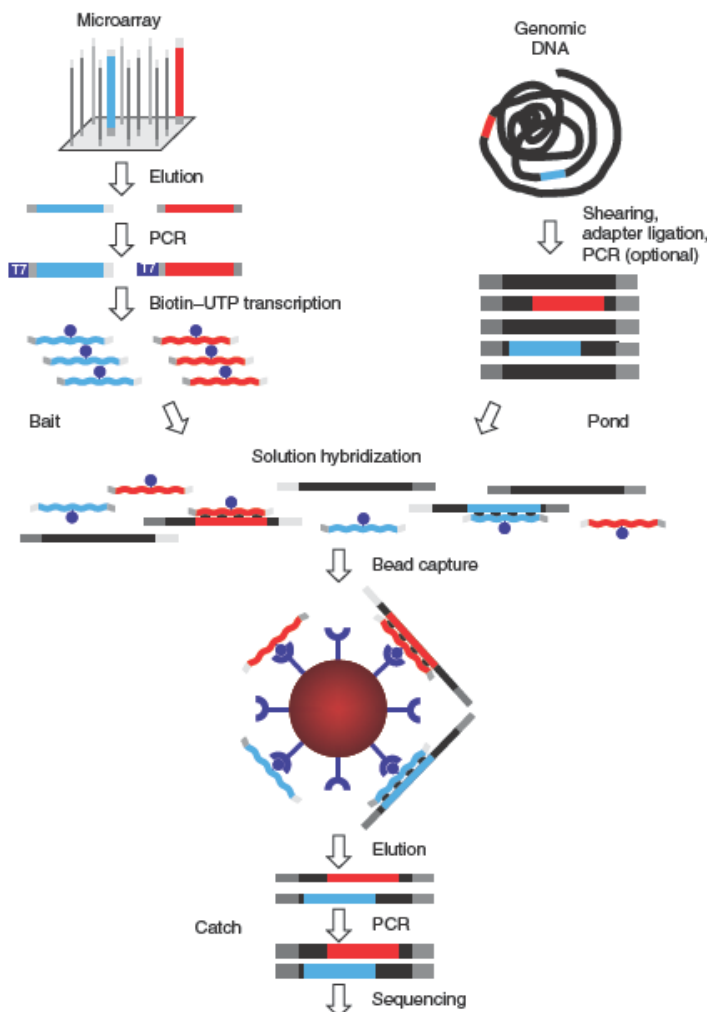
Merge

Output



PCR Droplet
37 microns - 26 pL

SureSelect (sequence capture)





Pilot Project

4 patients with NMD

Targeted enrichment

SureSelect

RainDance

NGS

Illumina
(72 bp reads)

SOLiD
(50 bp reads)

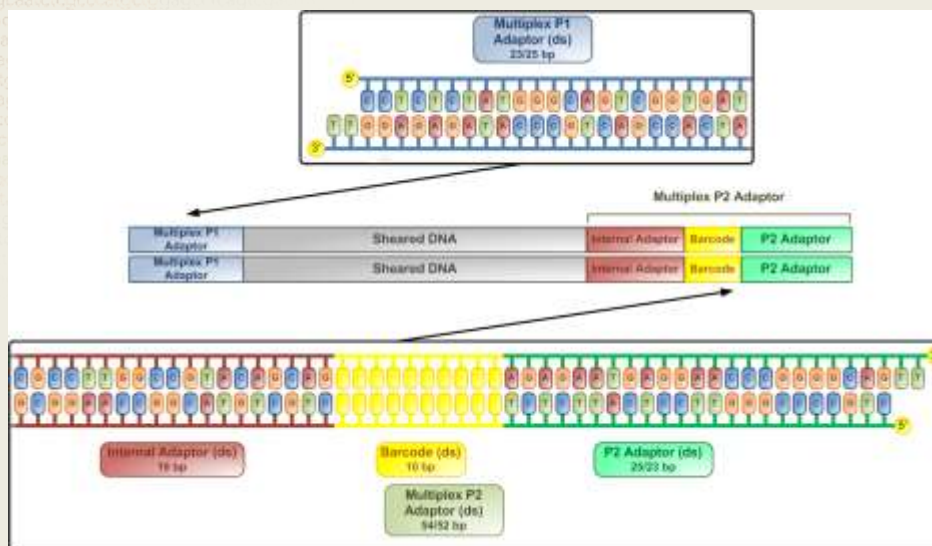
Illumina
(72 bp reads)

SOLiD
(50 bp reads)

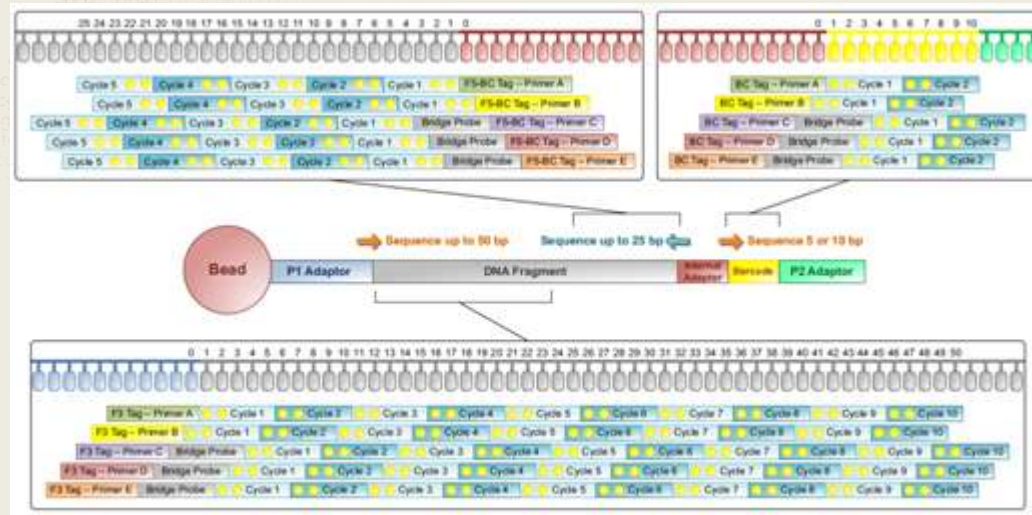


Bioinformatics

SOLiD barcoded library

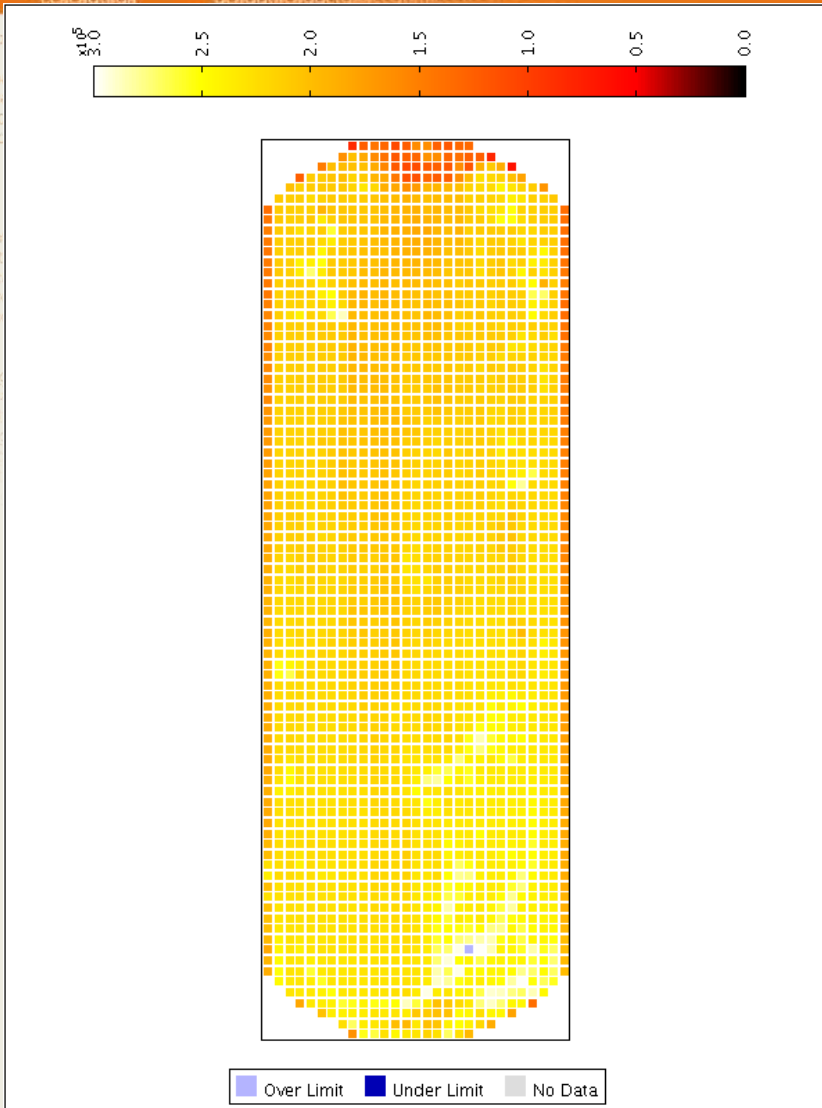


SOLiD dibase sequencing



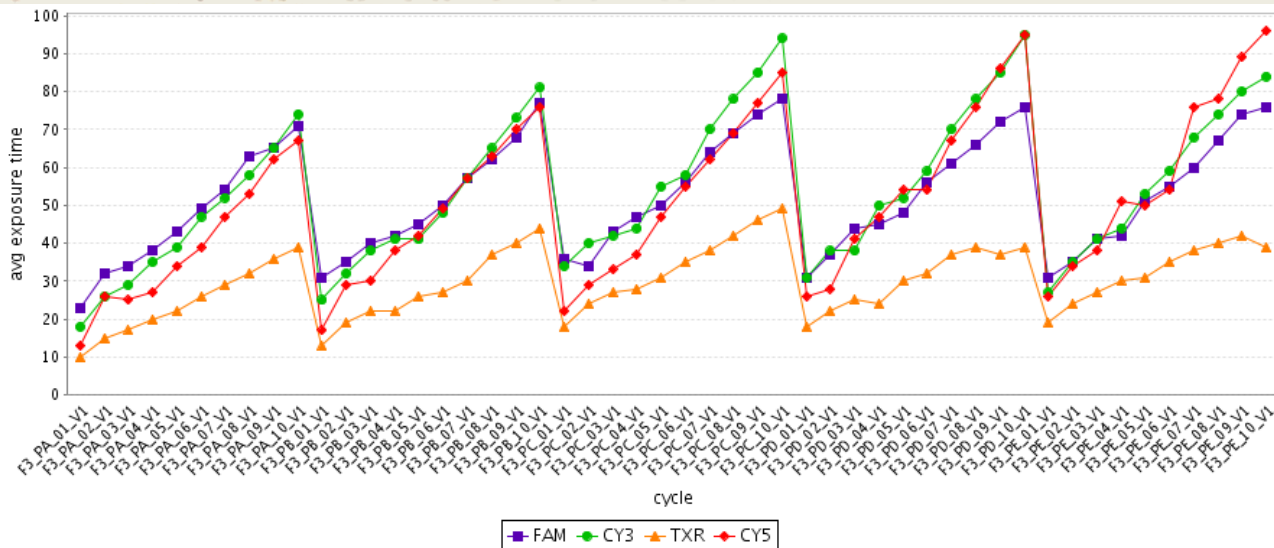


Heat map



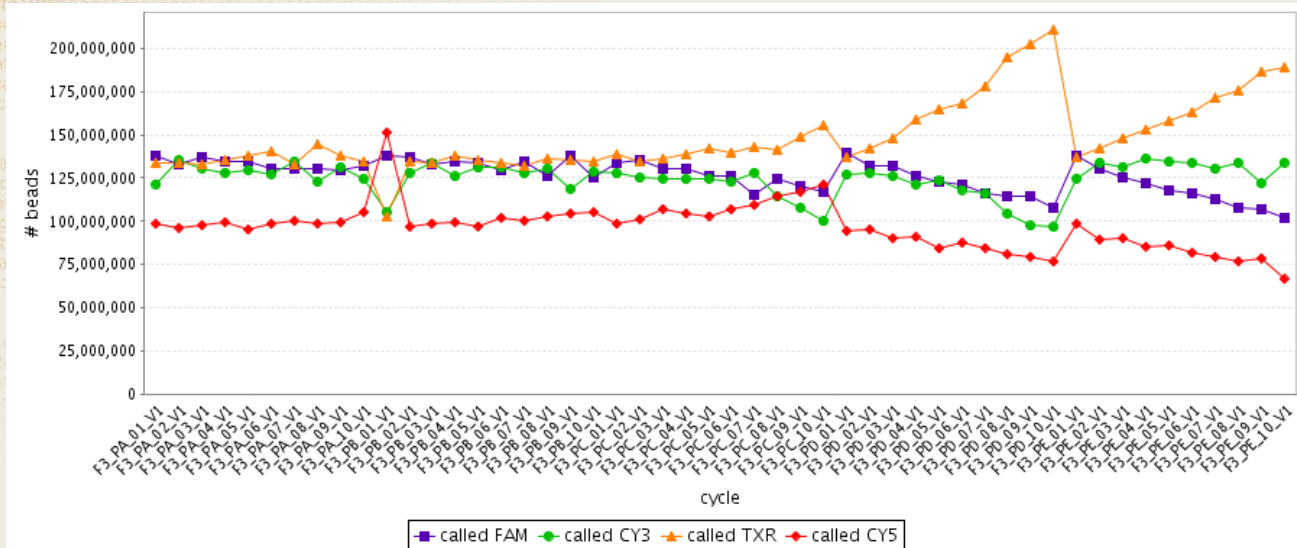
- Indicates bead density on the slide. In this particular case, the instrument detected 500 million beads

Exposure Time



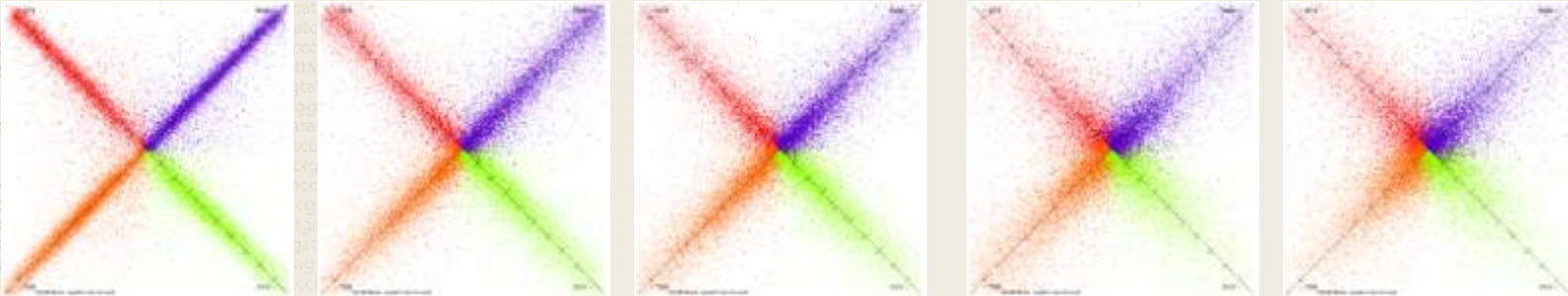
- Indicate intensity of four dyes. It tends to decrease as run progresses. Each of 5 primers have 10 ligation cycles. Intensity is lower at the end of the ligation cycles. Expectation is to have exposure time below 100 ms in the end on each primer.

Colour Balance



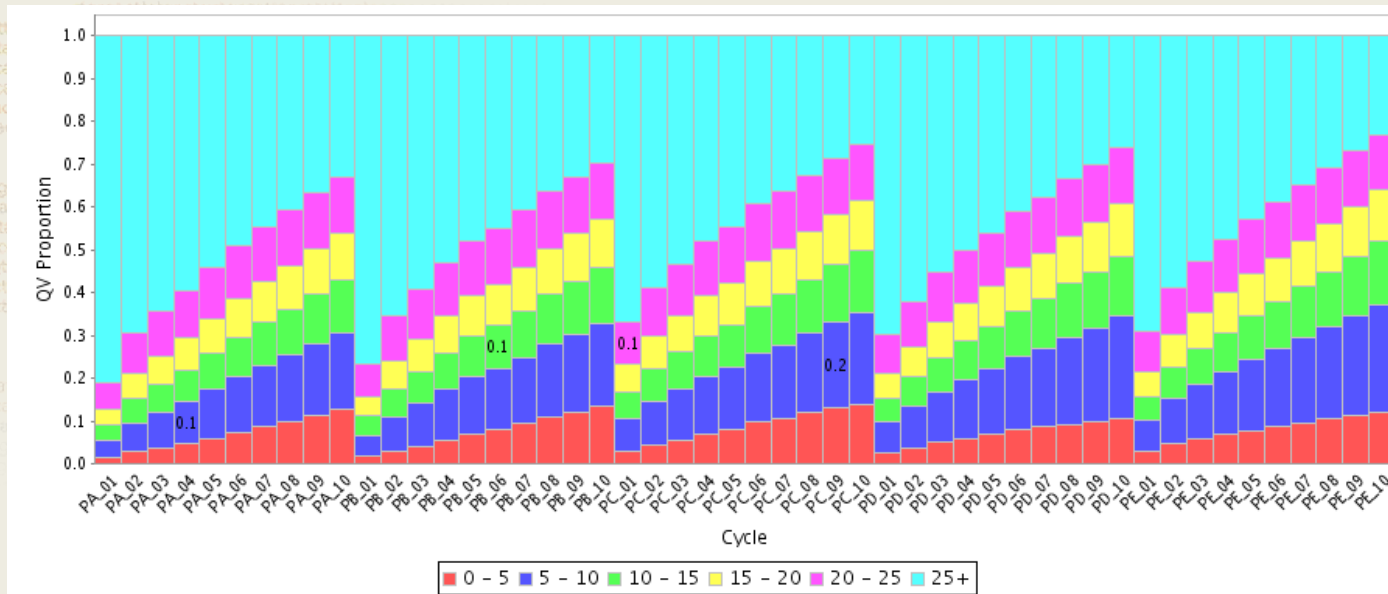
- It indicates how balanced the samples are regarding dibase composition. Skewed colour balance would result in low quality sequences. Not the case for these samples.

Satay Report



- Shown above for cycles 1, 3, 5, 8 and 10 of primer 1. It takes a random sampling of beads and measures the amount of monoclonal beads (on axis) and polyclonal beads (off axis), as well as dye intensity (lower intensity in the centre). As run progresses, we loose intensity and polyclonal beads become more evident.

Quality scores



- Expectation is to have 50% or more of Light Blue and Magenta – higher quality scores.

Preliminary Data



- Number of reads: between 3 and 100 million depending on the platform and density loaded.

Coverage: 819X on average

% bases with at least 2 reads: 95.5%

% bases with at least 10 reads: 88.4%

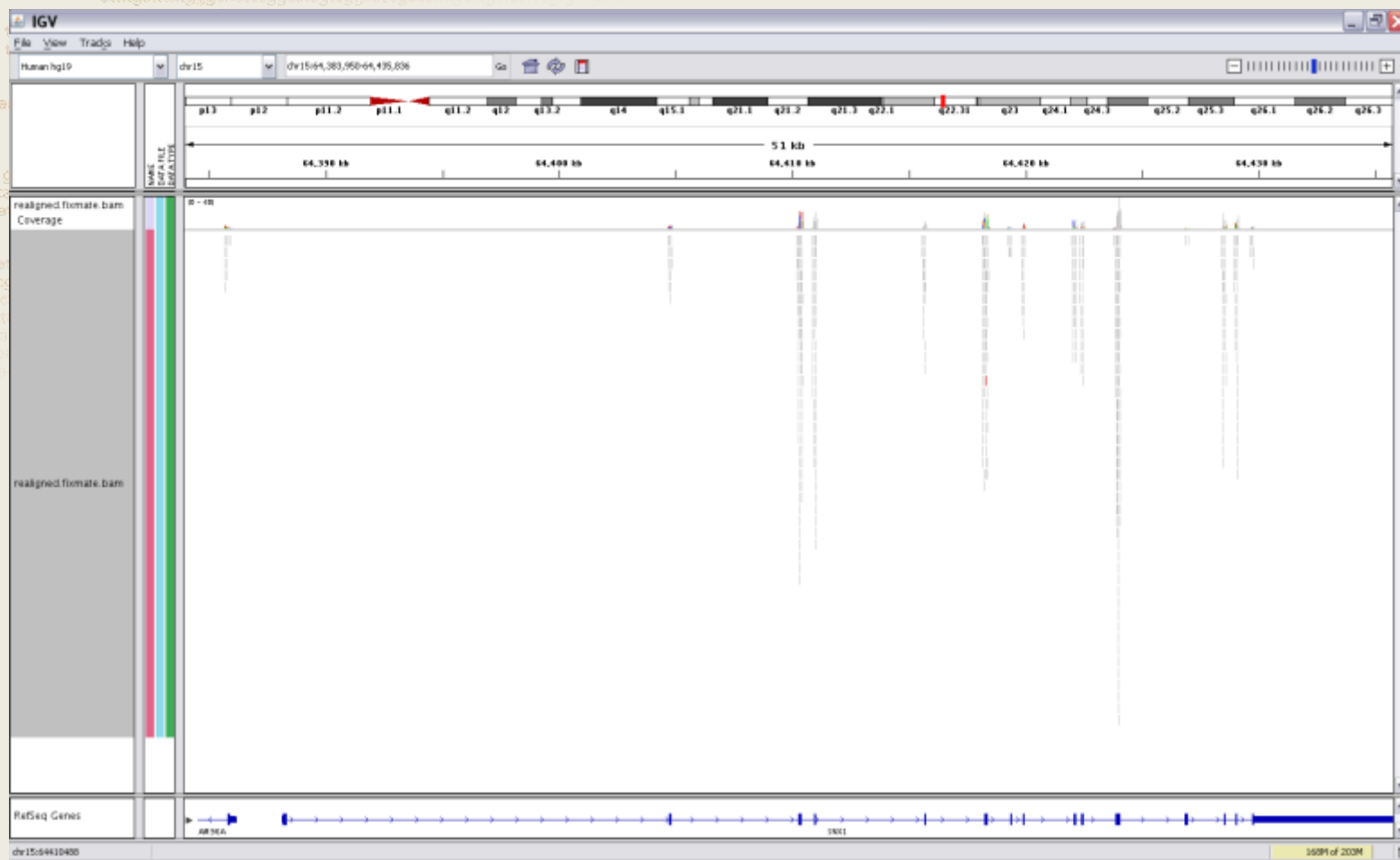
% bases with at least 20 reads: 82.1%

% bases with at least 30 reads: 75.4%

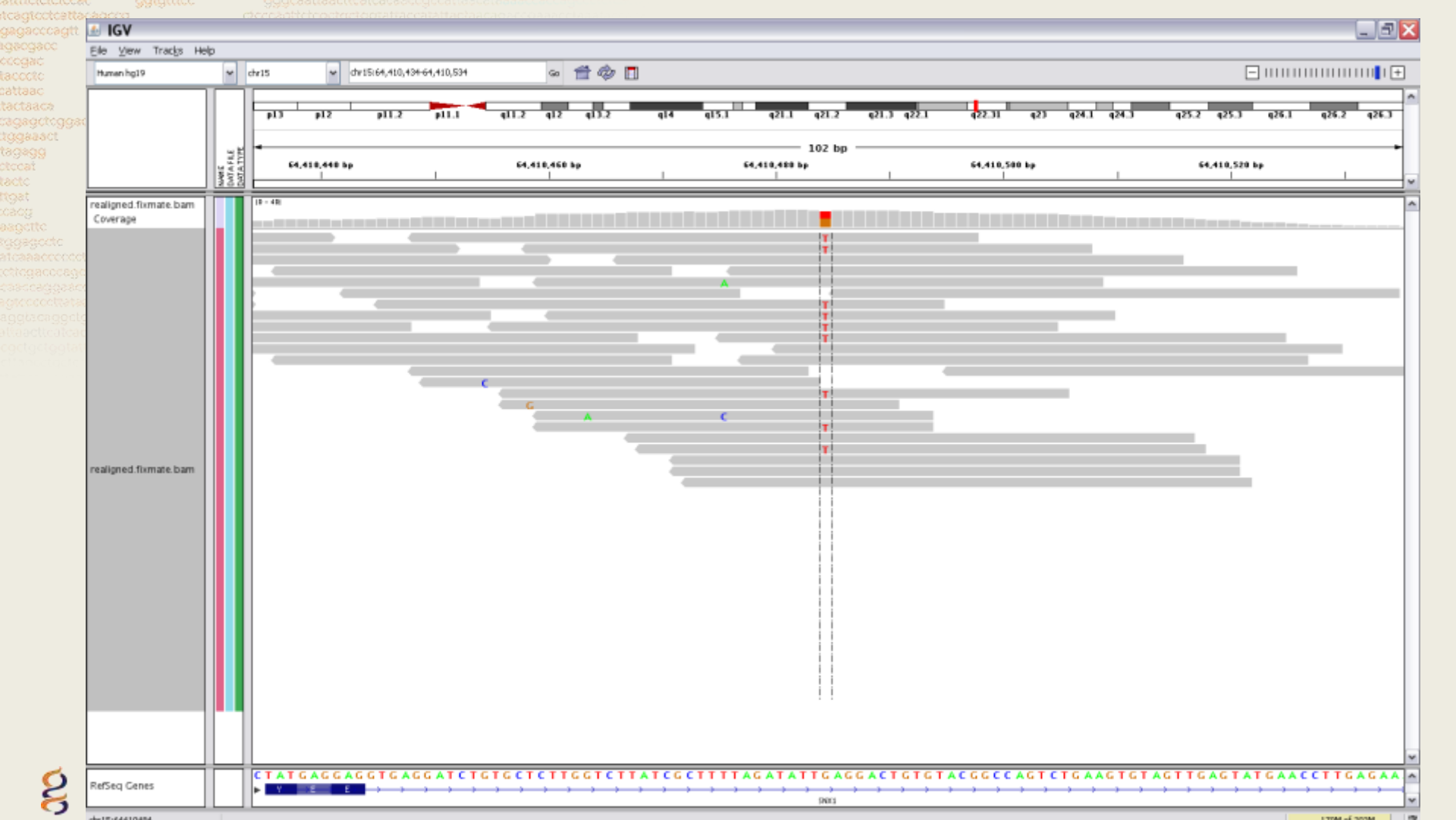
Amount of bases on baits: 416 Kb

Target territory: 911 Kb.

SNX1 gene



gatttttggcncocggcatg
 ttggtactgcccacgcttgcgtaataatctct
 aaacaataagctctgactctctcccccac
 gcccacatgctggagcctcagtc
 ctccaagatcaa
 aataccacaltc
 aacccctct
 ttataatt
 aggcctg
 tctatca
 ctggta
 tcaacg
 atgggat
 ctctctctcc
 tttctctctcc
 ttcagctctcattacacac
 gagagccagtt
 agccgacc
 ccgac
 taccctc
 cattaac
 tactaaca
 tagagctcggac
 tgggaaact
 tagagg
 ctccat
 tactc
 tgcg
 acacg
 aagcttc
 tggagctc
 atcaaaacccccc
 ctctogaccagc
 caaccaggaac
 agtcccctata
 aggtacngctc
 tctctatcaac
 tctctctctactgcttcccccagtaggggtaggag
 tctctctccaccaggggtttctctccatctagggg
 tcaigtctcattacagccgctcctactctactca
 accttatctcggggcatgagccggcatalaggga
 ttactgcccacgctctcgtataatctctctatg
 aatataagctctgactcttcccacacctctca
 atgctggagcctcagtcgacttagccattctca
 agtatcaaacccccctatttgtatgatcagctc
 cattcttcgacccagccggaggagggagccac
 ccaggaaacctctaggagacgacaaatcaaa
 gtccccctataattggctctccgacctagcctc
 gcaggfacaggctgaaccgtataccctccatagga



Data Analysis



Bioinformatic tools

In house - TCAG

Commercial tools for diagnostic labs –

SeqPilot

NextGene

Considerations

Exon coverage

Exon representation variation

Sequence accuracy



Conclusions



- Most likely will take up Agilent Sure select technology (but no real difference between Illumina and SOLiD technologies)
- Need to speed up bioinformatics

Acknowledgements



SickKids

Peter Ray

Jennifer Orr

TCAG

Steve Scherer

Sergio Pereira

Elango Cheran



